Statistics and Probability Letters 83 (2013) 2699–2702

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

A note on bootstrap confidence intervals for proportions

Weizhen Wang *

College of Applied Sciences, Beijing University of Technology, Beijing, 100124, PR China
Department of Mathematics and Statistics, Wright State University, Dayton, OH 45435, USA

ARTICLE INFO

Article history:
Received 14 June 2013
Received in revised form 5 September 2013
Accepted 5 September 2013
Available online 12 September 2013

Keywords:
Binomial distribution
Coverage probability
Multinomial distribution
Relative risk
Odds ratio

ABSTRACT

We first show that any $1-\alpha$ bootstrap percentile confidence interval for a proportion based on a binomial random variable has an infimum coverage probability zero for any sample size. This result is then extended to intervals for the difference, the relative risk and the odds ratio of two proportions as well as other types of bootstrap intervals.

1. Introduction

The bootstrap method (Efron, 1979; Hall, 1992; Efron and Tibshirani, 1993) has become widely used in statistical inference due to its accessibility. However, some researchers including Bickel and Freedman (1981), Athreya and Fukuchi (1994), Beran (1997), Bretagnolle (1983), Beran and Srivastava (1985), Dümbgen (1993), Romano (1988), Sriram (1993) and Beran (1997) have noticed that in some cases the bootstrap confidence interval is not consistent. Following Mantalos and Zografos (2008), an interval is said to be consistent if its coverage probability at each value of the parameter approaches the nominal level $1-\alpha$ as the sample size goes to infinity. Andrews (2000) discussed these cases in detail and concluded that they are all somewhat nonstandard. He then proposed a fairly standard counterexample (the normal distribution with an unknown nonnegative mean) that shows the bootstrap interval is not consistent.

In this note, we move one step further and focus on the infimum coverage probability of the bootstrap interval for inference on proportions. Two facts are worthy of mention. First, even when a confidence interval is consistent, there is still no guarantee that this interval based on a finite sample has a good chance to capture the parameter of interest. For example, the famous Wald interval for a single proportion is consistent, but it has a zero infimum coverage probability for any sample size. Following the recommendation by Agresti and Coull (1998), Brown et al. (2001) and others, the Wald interval should not be used in practice. In actual applications, only finite samples are available and the location of parameter is unknown. From the frequentist point of view, it is reasonable to protect against the worst case scenario, i.e., maintain the infimum coverage probability of a confidence interval to be at least or close to the nominal level for any given sample size. Second, inference on proportions is widely used in statistical applications and therefore, research on this topic can have a major impact on statistical practice.

In Section 2 it is shown that all bootstrap percentile intervals for a single proportion have a zero infimum coverage probability regardless of the sample size, the estimator being used and the confidence level and these results are extended to other types of bootstrap intervals. To the best of our knowledge, this seems to be the first study on the infimum coverage

* Correspondence to: Department of Mathematics and Statistics, Wright State University, Dayton, OH 45435, USA.
E-mail address: weizhen.wang@wright.edu.

0167-7152/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.spl.2013.09.005

probability on bootstrap intervals. Due to the availability of exact admissible confidence intervals (e.g. Blyth and Still, 1983; Casella, 1986; Wang, 2013), bootstrap intervals for a single proportion might not be preferred in applications. However, for more complicated cases (discussed below) bootstrap intervals can be of interest. In Section 3, bootstrap intervals for the difference, the relative risk and the odds ratio of two independent proportions are shown to have a zero infimum coverage probability as well. Similar results are presented in Section 4 for the multinomial distribution including the comparison of two dependent proportions. Therefore, the zero infimum coverage probability for a bootstrap interval is not rare. It is known that the Wald interval collapses to the single number 1 if no failure is observed. This may occur when the true proportion is close to one, which leads to a zero infimum coverage for the Wald interval. The same logic can be applied to bootstrap intervals for proportions.

2. Bootstrap interval for a proportion p

Suppose $\underline{X} = \{X_i\}_{i=1}^n$ is an i.i.d. sample from a Bernoulli population with a probability of success p. Then $X = \sum_{i=1}^n X_i$ is a binomial random variable with n trials and a probability of success p, denoted by $Bin(n, p)$. Let $\underline{X}^B = \{X_i^B\}_{i=1}^n$ be a bootstrap sample (i.e., a simple random sample with replacement) of size n from $\underline{X}$. Then $X^B = \sum_{i=1}^n X_i^B$ follows $Bin(n, p(X))$ with $p(X) = X/n$ for each observed X. For any estimator $\hat{p}(\underline{X}^B)$, consider its $1-\alpha$ bootstrap percentile confidence interval for p based on m bootstrap samples,

$$C(X, \hat{p}(\underline{X}^B), m) = [L(X, \hat{p}(\underline{X}^B), m), U(X, \hat{p}(\underline{X}^B), m)]$$

with confidence limits L and U. That is, for each bootstrap sample $\underline{X}_j^B$ from $\underline{X}$ for $j = 1, \ldots, m$, we compute $\hat{p}(\underline{X}_j^B)$ and obtain $A = \{\hat{p}(\underline{X}_j^B)\}_{j=1}^m$. Then $L(X, \hat{p}(\underline{X}^B), m)$ and $U(X, p(\underline{X}^B), m)$ are equal to the $(\alpha/2)$th and $(1-\alpha/2)$th percentiles of A, respectively. For simplicity, rewrite $C(X, \hat{p}(\underline{X}^B), m)$ as $C(X, \hat{p})$.

Example 1. Suppose we observe $9(=X)$ successes from $30(=n)$ independent and identical Bernoulli trials with an unknown p. Suppose a 95% bootstrap interval for p is to be computed based on the estimator

$$\hat{p} = \frac{X}{30}\left(\frac{30}{30+1.96^2}\right) + \frac{1}{2}\left(\frac{1.96^2}{30+1.96^2}\right),$$

which is the center of the 95% Wilson interval (1927). The observed sample $\underline{X}$ consists of nine 1's and twenty one 0's. We take $10000(=m)$ bootstrap samples $\underline{X}_j^B$, $j = 1, \ldots, 10000$ from $\underline{X}$, compute $\hat{p}$ for each j and store all bootstrap estimates in set A. The 2.5th and 97.5th percentiles of A are the observed bootstrap percentile confidence limits L and U. If repeat this process, the resultant interval limits would almost certainly be a little different. On the other hand, if all $30(=X)$ trials are successes, then all 10000 elements of A are identical and equal to $c = \frac{30+1.96^2/2}{30+1.96^2}$, in which case $L = U = c$. Therefore, it is possible that the bootstrap percentile interval reduces to a single point for certain sample points. □

Lemma 1. *For any $\alpha \in (0, 1)$, $n \geq 1$ and $m \geq 1$, and for any estimator $\hat{p}$, let $C(X, \hat{p})$ be its $1-\alpha$ bootstrap percentile interval for p. Then the coverage probability of $C(X, \hat{p})$, given by*

$$f(p) = P_p(p \in C(X, \hat{p})), \tag{1}$$

has an infimum zero.

Proof. When $X = 0$, it is clear that $\underline{X}^B$ is always a sequence of n zeros. Therefore, A is a set of m constants c, where $c = \hat{p}(0, \ldots, 0)$. This leads to $L(0, \hat{p}) = c = U(0, \hat{p})$, and $C(0, \hat{p})$ reduces to a single point c. Pick a sequence of proportions $p_i(>0)$ that are not equal to c such that the sequence approaches zero. Then

$$\inf_{p\in[0,1]} f(p) \leq \lim_{i\to+\infty} P_{p_i}(p_i \in C(X, \hat{p})) \leq \lim_{i\to+\infty} [1 - P_{p_i}(X = 0)] = \lim_{i\to+\infty} [1 - (1-p_i)^n] = 0. \quad \square \tag{2}$$

In particular, the two bootstrap intervals based on the sample proportion ($\hat{p}_S = X^B/n$) and the center of the Wilson interval have a zero infimum coverage probability.

Remark 1. It is easy to see that for other types of bootstrap intervals, such as the basic bootstrap interval and the bootstrap percentile-t interval (Engel, 2010; Mantalos and Zografos, 2008), that when $X = 0$, there is no variation in the sample of $\underline{X} = \{X_i\}_{i=1}^n$ and these bootstrap intervals all reduce to a single point. Therefore, (2) can be applied and Lemma 1 is also true for these intervals. Some variations of this simple fact will be utilized in the following sections to establish similar results for more interesting and complicated cases. □

Remark 2. There is another type of bootstrap interval: the parametric bootstrap interval (Efron, 1985), where the maximum likelihood estimator (the sample proportion $\hat{p}_S$) is used to estimate p and then parametric bootstrap samples are collected by sampling from the Bernoulli distribution with $p = \hat{p}_S$. When $X = 0$, then $\hat{p}_S = 0$ and any parametric bootstrap sample is a sequence of 0's without any variation. Thus the parametric bootstrap interval is equal to [0, 0] and (2) can be applied, which again leads to a zero infimum coverage probability for this interval. □

3. The comparison of two independent proportions

Consider the case of observing two independent $X \sim Bin(n_1, p_1)$ and $Y \sim Bin(n_2, p_2)$. We are interested in bootstrap percentile interval for a parameter $\theta = g(p_1, p_2)$ for a given function g. The three most common examples of $g(p_1, p_2)$ for comparing two proportions are $\theta_{di} \overset{def}{=} p_1 - p_2$, $\theta_{ri} \overset{def}{=} p_1/p_2$ and $\theta_{oi} \overset{def}{=} p_1(1-p_2)/(p_2(1-p_1))$ corresponding to the difference, the relative risk and the odds ratio, respectively. Some examples of using bootstrap intervals to estimate θ_{di} and θ_{oi} and individual p_i's under a monotone constraint are given in Lin et al. (2009), Parzen et al. (2002) and Li et al. (2010).

To estimate a parameter of interest $\theta = g(p_1, p_2)$ with an observation (X, Y), a bootstrap percentile interval is derived based on a certain point estimator $\hat{\theta}(X, Y)$ and m bootstrap samples, denoted by $C(X, Y, \hat{\theta}, m)$. When $(X, Y) = (0, n_2)$, similar to the previous section, the bootstrap sample $\underline{X}^B$ is a sequence of n_1 0's and $\underline{Y}^B$ is a sequence of n_2 1's. Thus $C(0, n_2, \hat{\theta}, m)$ reduces to a single point $d = \hat{\theta}(0, n_2)$. Choose a sequence $p_{1i}(>0)$ approaching 0 and a sequence $p_{2i}(<1)$ approaching 1, and $g(p_{1i}, p_{2i}) \neq d$, then the infimum coverage satisfies

$$\inf_{0 \le p_1, p_2 \le 1} P_{p_1,p_2}(\theta \in C(X, Y, \hat{\theta}, m)) \le \lim_{i \to +\infty} [1 - P_{p_{1i},p_{2i}}(X = 0, Y = n_2)] = 0,$$

which is similar to (2). To summarize the above discussion, we have the following lemma.

Lemma 2. *For any $\alpha \in (0, 1)$, $n_1 \ge 1$, $n_2 \ge 1$, $m \ge 1$, and for any parameter of interest $\theta = g(p_1, p_2)$ with an estimator $\hat{\theta}$, let $C(X, Y, \hat{\theta}, m)$ be its $1 - \alpha$ bootstrap percentile interval for θ. Then the coverage probability of $C(X, Y, \hat{\theta}, m)$, given by*

$$f(p_1, p_2) = P_{p_1,p_2}(\theta \in C(X, Y, \hat{\theta}, m)), \tag{3}$$

has an infimum zero.

Remark 3. Similar to Remarks 1 and 2, the zero infimum coverage probability also occurs for the basic bootstrap interval, the bootstrap percentile-t interval, and the parametric bootstrap interval based on the maximum likelihood estimator. □

4. The case of the multinomial distribution

Suppose $\underline{X} = (N_1, \ldots, N_k)$ is observed from a multinomial experiment with n trials and probabilities $p_1, \ldots, p_k$ for k possible outcomes in each trial, and N_j is the number of trials in which the jth outcome occurs for $j = 1, \ldots, k$. We are interested in a bootstrap interval for some parameter of interest $\eta = h(p_1, \ldots, p_k)$ for a given function h based on an estimator $\hat{\eta}(\underline{X})$. The matched pair experiment (Agresti, 2002, p. 410) is a special case with $k = 4$ where $p_{11} = p_1$, $p_{12} = p_2$, $p_{21} = p_3$ and $p_{22} = p_4$. The difference $p_{12} - p_{21}$, the relative risk $(p_{11} + p_{12})/(p_{11} + p_{21})$ and the odds ratio $p_{11}p_{22}/(p_{12}p_{21})$ are often of interest when comparing a treatment with a control when there is a large degree of variation among subjects. Glaz and Sison (1999) gave another example of using parametric bootstrap interval to estimate the minimum and the maximum of p_j's.

Lemma 3. *For any $\alpha \in (0, 1)$, $n \ge 1$, $m \ge 1$, and for any parameter of interest $\eta = h(p_1, \ldots, p_k)$ with an estimator $\hat{\eta}(\underline{X})$, let $C(\underline{X}, \hat{\eta}, m)$ be its $1 - \alpha$ bootstrap percentile interval for θ. Then the coverage probability of $C(X, Y, \hat{\theta}, m)$, given by*

$$f(p_1, \ldots, p_k) = P_{p_1,\ldots,p_k}(\eta \in C(\underline{X}, \hat{\eta}, m)), \tag{4}$$

has an infimum zero.

The proof is similar to Lemma 2 and is omitted. The same holds for other types of bootstrap intervals.

5. Summary

In this note, we have shown that bootstrap confidence intervals for proportions always miss the target parameters with a probability close to one for certain value(s) of the parameter for any sample size and using any estimator. Therefore, while bootstrap intervals for proportions are easily accessible, it is important for researchers to realize the risk of using these interval. As shown in the proof of lemmas, if there exists a sample point with a probability mass close to one at some values of the parameter (or distributions), and the bootstrap sample at this point does not have any variation, then the infimum coverage probability of bootstrap interval is close to zero. This may also occur for a nonparametric distribution family.

Acknowledgments

The author is grateful to an anonymous referee and Professor Thaddeus Tarpey for their helpful comments.

References

Agresti, A., 2002. Categorical Data Analysis, second ed. John Wiley & Sons, Inc.
Agresti, A., Coull, B.A., 1998. Approximate is better than "exact" for interval estimation of binomial proportions. The American Statistician 52, 119–126.
Andrews, D., 2000. Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. Econometrica 68, 399–405.
Athreya, K.B., Fukuchi, J., 1994. Bootstrapping extremes of i.i.d. random variables. Proceedings of the Conference on Extreme Value Theory NIST.

Beran, R., 1997. Diagnosing bootstrap success. Annals of the Institute of Statistical Mathematics 49, 1–24.
Beran, R., Srivastava, M.S., 1985. Bootstrap tests and confidence regions for functions of a covariance matrix. Annals of Statistics 13, 95–115.
Bickel, P.J., Freedman, D., 1981. Some asymptotic theory for the bootstrap. Annals of Statistics 9, 1196–1217.
Blyth, C.R., Still, H.A., 1983. Binomial confidence intervals. Journal of the American Statistical Association 78, 108–116.
Bretagnolle, J., 1983. Lois limites du bootstrap de certaines fonctionnelles. Annales de L'institut Henri Poincarcé, Sec. B 19, 281–296.
Brown, L.D., Cai, T.T., DasGupta, A., 2001. Interval estimation for a binomial proportion. Statistical Sciences 16, 101–133.
Casella, G., 1986. Refining binomial confidence intervals. The Canadian Journal of Statistics 14, 113–129.
Dümbgen, L., 1993. On nondifferentiable functions and the bootstrap. Probability Theory and Related Fields 95, 125–140.
Efron, B., 1979. Bootstrap methods: another look at the jackknife. Annals of Statistics 7, 1–26.
Efron, B., 1985. Bootstrap confidence intervals for a class of parametric problems. Biometrika 72, 45–58.
Efron, B., Tibshirani, R., 1993. An Introduction to the Bootstrap. Chapman and Hall, New York.
Engel, J., 2010. On teaching bootstrap confidence intervals, in: Proceedings of the Eighth International Conference on Teaching Statistics, ICOSTS8, Ljubljana, Slovenia.
Glaz, J., Sison, C.P., 1999. Simultaneous confidence intervals for multinomial proportions. Journal of Statistical Planning and Inference 82, 251–262.
Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer-Verlag, New York.
Li, Z., Taylor, J.M.G., Nan, B., 2010. Construction of confidence intervals and regions for ordered binomial probabilities. The American Statistician 64, 291–298.
Lin, Y., Newcombe, R.G., Lipsitz, S., Carter, R.E., 2009. Fully specified bootstrap confidence intervals for the difference of two independent binomial proportions based on the median unbiased estimator. Statistics in Medicine 28, 2876–2890.
Mantalos, P., Zografos, K., 2008. Interval estimation for a binomial proportion: a bootstrap approach. Journal of Statistical Computation and Simulation 78, 1251–1265.
Parzen, M., Lipsitz, S., Ibrahim, J., Klar, N., 2002. An estimate of the odds ratio that always exists. Journal of Computational and Graphical Statistics 11, 420–436.
Romano, J., 1988. Bootstrapping the mode. Annals of the Institute of Statistical Mathematics 40, 565–586.
Sriram, T.N., 1993. Invalidity of the bootstrap for critical branching processes with immigration. Annals of Statistics 22, 1013–1023.
Wang, W., 2013. An iterative construction of confidence intervals for a proportion. Statistica Sinica (in press).
Wilson, E.B., 1927. Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association 22, 209–212.